TRANSMITTED HEREWITH ARE 21 COPIES OF AN AFCRL REPORT. IT IS REQUESTED THAT THE AD NUMBER BE PLACED ON ONE COPY OF THIS REPORT AND THE COPY WITH THE AD NUMBER BE RETURNED TO AFCRL-CRXRA
-[redacted]-, STOP 39, L G HANSCOM FLD, MASS.

AD-401 920

Research Note

Extraction of Pitch From the Trachea

HAROLD C. PORTER

COMMUNICATION SCIENCES LABORATORY PROJECT 4610

AIR FORCE CAMBRIDGE RESEARCH LABORATORIES, OFFICE OF AEROSPACE RESEARCH, UNITED STATES AIR FORCE, L.G. HANSCOM FIELD, MASS.

Requests for additional copies by Agencies of the Department of Defense, their contractors, and other government agencies should be directed to the:

Armed Services Technical Information Agency
Arlington Hall Station
Arlington 12, Virginia

Department of Defense contractors must be established for ASTIA services, or have their 'need-to-know' certified by the cognizant military agency of their project or contract.

All other persons and organizations should apply to the:

U. S. DEPARTMENT OF COMMERCE
OFFICE OF TECHNICAL SERVICES,
WASHINGTON 25, D. C.

Abstract

For decades, speech researchers have recognized that accurate pitch information is prerequisite to the synthesis or reconstitution of speech that has been broken into its basic parts for any reason, such as communication over restricted bandwidth media.

In a series of experiments investigating tracheal pitch extraction according to a methodology suggested in a Japanese paper, the Speech Research Branch, AFCRL Communication Sciences Laboratory succeeded in achieving reasonably accurate real-time pitch extraction. The method has a limitation in that the talker must wear a special appliance.

Foreword

The extraction of voicing information from a speech signal is generally held to be a difficult task, especially if the signal is band-limited, and the evaluation of devices that perform this function—so-called "pitch extractors"—is hampered by the experimenter's knowing neither just how many glottal excitation pulses were produced in the original speech nor just when each of them occurred. If either or both of these could be easily measured it would be extremely helpful.

In 1960, Sugimoto and Hiki[1] reported on a technique of closely coupling an electroacoustic transducer to the skin covering the anterior aspect of the trachea at a level below the vocal cords and near the clavicle, from which point the transducer output (1) had a good level, (2) was an accurate representation of the glottal pulsation, (3) was stable over changes in vocal tract configuration and pronunciation, and (4) did not obstruct the speech of the talker.* Although such a technique would be difficult to incorporate into a system design, it could be extremely valuable in the laboratory for deriving basic information about the voicing process as well as data that could be applied in specific evaluation problems. Accordingly, it was decided to investigate how well such a technique might work with transducers readily available, and whether the results might be useful in our research program.

*These results were implicit in some AFCRC-sponsored experiments performed at Northeastern University, reported in Final Report under ARDC Contract AF19(604)-3465: Speech Analysis, AFCRC-TR-59-187, 31 August 1959 (pages 18 to 21).

Application of this technique involves choosing: (1) a transducer, (2) a method of coupling the transducer to the throat of the talker so as to maximize pickup from the throat wall and minimize pickup from the air outside, and (3) a position on the throat where the transducer should be placed. The mechanical configuration and the electrical output of the transducer that result from these choices must not only satisfy certain requirements for acceptability but must optimize the following characteristics.

First, the electrical output must exhibit an adequate signal-to-noise ratio. This is indicated when the glottal periodicity is the predominant characteristic of the waveform. Signal, then, is transducer output that indicates glottal pulsation and is picked up through the tracheal wall, whereas noise is anything in the transducer output that obscures that indication. On this basis the following are classed as noise: (1) external speech signals, (2) any external noise, and (3) unvoiced speech sounds or the fricative parts of voiced speech sounds picked up through the wall of the throat. In one sense there is a fourth kind of noise, namely, those "harmonics" of the voicing "frequency" that are not necessary to form the glottal pulse and may actually tend to mask it. On the other hand, it may be that these harmonics are not part of the primary excitation but are instead "feedback" from the vocal tract through the glottis. A close inspection of the amplitudes of these harmonics throughout the voicing cycle may tell when the vocal folds are open or closed, in which case these harmonics might better be considered as signal.

Second, it is highly desirable for the signal peaks to be relatively invariant in amplitude for all vocal tract configurations used during voicing; any variations that do appear in the output should not be attributable to the transducer, the method of coupling, or the positioning of the transducer on the throat. Even when the peak amplitudes remain relatively constant, however, the waveform underneath the peaks may not. It is an experimental fact that the waveform changes with changes in vocal tract configuration during voicing in the lower registers. In this case, extraneous harmonics are never completely absent from the transducer output.

Third, the transducer must be so light, and the method of securing it at the throat of the talker so comfortable, that perfectly normal speech results. Further, the speaker must be allowed reasonable freedom of bodily movement such that neither the coupling nor the proper positioning of the transducer is disturbed.

Fourth, the transducer must have a flat response-frequency characteristic and a linear phase characteristic from below 50 cps to above 500 cps. According to Sugimoto and Hiki,[2] the transfer function through the tracheal wall has a low-pass characteristic. Their average curve for amplitude as a function of frequency begins to drop at about 400 cps and has a slope of 36 db/octave. This curve is apparently based on data obtained from adult voices. If the lighter wall of a child's trachea is considered, it could be conjectured that the cutoff frequency would be even higher,

and hence that a general-purpose transducer would have to have a flat response and good waveform transfer to beyond 500 cps.

Mr. Porter's work confirms the earlier Japanese results and indicates that his realization of the technique reasonably satisfies the four requirements just discussed. It must nevertheless be kept clearly in mind that the usefulness of this technique is predicated on the assumption that the transfer function between glottis and transducer is essentially constant and consequently that the time delay between the glottal event and its indication in the output is likewise constant. If time relations are thus preserved, the time difference between any two adjacent pulses seen in the output is a valid measure of the difference between the corresponding adjacent glottal excitations. In other words, the measurements are accurate in a relative but not an absolute sense, although the difference is of course extremely small, being some fraction of a millisecond.

We are now at the point where we think the throat pickup gives a highly accurate indication of glottal excitations, but evaluation of the technique has so far been only qualitative. Comparison of the waveform picked up at a talker's throat and the acoustic waveform in front of his mouth shows more pulses in the former than in the latter. There is no reason to assume that these extra pulses are not valid. They occur, for example, during closures in voiced stops and nasals where the external acoustic signal is almost nonexistent and the throat signal is very strong. Consequently, the external signal is not a completely valid criterion of the performance of the throat pickup. The only real test is to compare the output of the throat pickup with the vocal cord activity seen by direct examination. This we have not yet done.

WEIANT WATHEN-DUNN, Chief,
Speech Research Branch, AFCRL
Communication Sciences Laboratory

Illustrations

Extraction of Pitch From the Trachea

1. METHOD

This series of pitch extraction experiments was based on the use of a transducer pressed firmly against the throat over the trachea. Regardless of the transducer used, the output signal contained obvious periodicity when observed on an oscilloscope. It was easy to correlate this periodicity with that of the simultaneous output of a standard microphone energized by the associated normal speech of the subject under test (Fig. 1). To analyze the tracheal signal electronically, it

FIG. 1. Comparison of waveform periodicity in acoustic tracheal signals of the same speech utterance: upper, normal external acoustic signal; lower, signal picked up at the trachea.

Submitted for publication 17 October 1962

was found advisable to filter, center clip, and limit the output of the tracheal transducer. This produced a train of pulses of uniform amplitude free of extraneous data. The pulses could then be counted during a standard time interval to get the average frequency of the pitch, or the time between adjacent pulses could be measured against a standard frequency to get the periodicity. Both methods were tried, but the latter was adopted because it is more exact.

The entire lower area of the throat was explored to find the point yielding the most usable signal. The best pickup was consistently obtained from a central point over the trachea just below the larynx (Fig. 2). Good signals were also

FIG. 2. Position of the tracheal transducer that consistently gave the best pickup.

obtained from points just above the clavicle left and right of the trachea. The center position was adopted as the standard for all further tests.

The standardized evaluation procedure involved simultaneous magnetic-tape recording of the tracheal signal and the signal from a high-quality microphone placed for normal voice pickup. The recorded signals were then compared, either on an oscilloscope, or by detailed examination of a 35-mm film of the oscilloscope display.

2. INSTRUMENTATION

The experimental instrumentation included transducers, electronic processing equipment, and display and recording devices.

2.1 Transducers

A transducer is necessary to transform the slight physical vibrations of the outer trachea surface into a usable electrical signal. The several that were tried will be discussed in order.

2.1.1—In our initial test an Altec 633A high-quality dynamic microphone was coupled to the throat over the trachea surface by a ring of sponge rubber enclosing an air space between the microphone diaphragm and the throat. Because the sides of the effective coupler were stiff and the dimensions were a small fraction of any wavelength under consideration, the pressure changes at the tracheal end were faithfully duplicated at the diaphragm. Without filtering or any processing other than linear amplification, the fundamental pitch period was readily discernible at the output of the microphone (Fig. 3). After tests with this microphone had shown

FIG. 3. Fundamental periodicity derived from an Altec 633A microphone: upper, normal external acoustic signal; lower, tracheal signal without further processing.

the feasibility of this approach, a search was started for a smaller, lighter transducer that could be worn without discomfort.

2.1.2—A crystal vibration pickup of the type commonly used on musical instruments was mounted on an elastic throat band and oriented so that the tracheal vibrations would actuate it. Constant coupling between the trachea and the pickup proved impossible to maintain. In addition, the pickup had a very poor frequency-response characteristic, introducing intolerable amounts of frequency distortion. This pickup was discarded after a few tests.

2.1.3—As a third approach we tried a Western Electric 640AA condenser microphone measuring about 1 in. in length by 1 in. in diameter, normally the endpiece of a rather large preamplifier unit. A series of tests showed that the microphone alone could be operated at the end of several feet of miniature coaxial cable without

degrading the frequencies of interest, enabling us to position the microphone against the throat with a neckband and locate the amplifier at some convenient point. As with the dynamic microphone, the condenser unit was initially coupled to the throat by a sponge rubber ring. Pressing the face of the unit lightly against the skin to eliminate any air space other than the small gap between the grille and the diaphragm gave better results. This could also have been the case for the dynamic

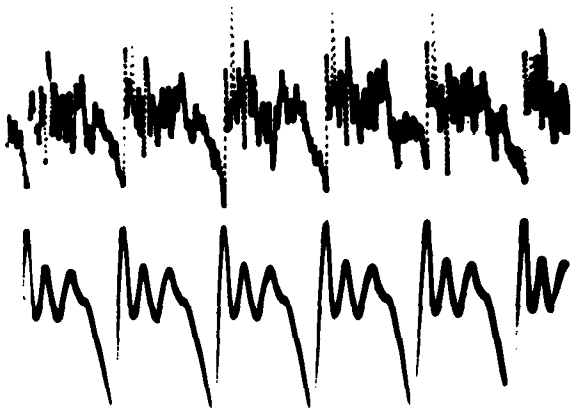

FIG. 4. Fundamental periodicity derived from a Western Electric 640AA microphone: upper, normal external acoustic signal; lower, tracheal signal without further processing.

FIG. 5. Pickup modifications: upper, stethoscope coupled to WE 640AA on RA-1095 preamplifier; lower, Altec 21C with rubber grommet coupler mounted on extension cable.

microphone except that its construction would not permit such tight coupling. Results of the tests with the condenser unit (Fig. 4) were slightly superior to those obtained with the larger unit, with the added advantage that we now had a device that could be comfortably worn by the talker for short periods. Concern for the comfort of the talker led us to explore the possibilities of picking up the tracheal vibrations by means of a stethoscope arrangement coupled to the condenser microphone by a small plastic tube. A small light brass horn 1 in. in diameter was

fabricated and coupled to the face of the condenser unit by about 2 ft of 1/4-in. I.D. plastic tubing (upper assembly, Fig. 5). Unfortunately, 2 ft exceeds one wavelength for the higher frequencies of interest, and standing waves introduced undesirable peaks and nulls into our signal. The tubing was then treated as the acoustic analog of an electrical transmission line, and the acoustic analog of a matching stub in the form of a short length of tubing with a variable port was inserted into the tubing near the microphone. By varying the size of the port it was possible to restrict the amplitude deviations in the range of interest (50 to 500 cps) to less than 3 db (Fig. 6). The input signal, derived from a sweep frequency

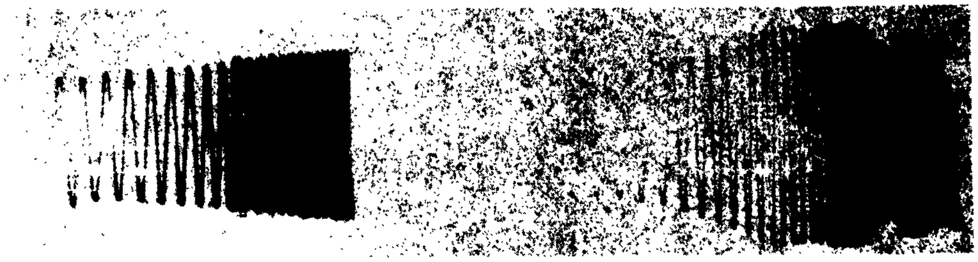

FIG. 6. Response-frequency characteristic of stethoscope pickup: left, voltage into driving earphone; right, voltage out of RA-1095 amplifier.

generator driving a PDR-10 high-quality headphone connected to a standard 6-cc acoustic coupler, did not have a uniform frequency-response characteristic. As a further refinement, the length of the "transmission line" was adjusted so that the acoustic delay of the tracheal output was approximately equal to the delay through the vocal tract. This enabled us to observe the correct phase relationship between the two signals. In general, for pitch extraction purposes, the results of this method almost equalled those of the direct condenser-microphone pickup. Two minor annoyances were a tendency of the plastic tubing to buckle when bent and the leakage of loud extraneous noises into the system through the thin flexible walls of the tubing.

3.1.4—At this point an Altec 21C microphone with its associated preamplifier became available to us. This condenser microphone is 1/2 in. long and approximately 5/8 in. in diameter. It is very light, and is normally used as an integral part of a Model 150 preamplifier, being affixed to the end of the wandlike unit. It was determined by experiment that for our purposes, without apparent degradation of the signals of interest to us, the condenser unit could be operated physically remote from the preamplifier, connected to it by about 2 ft of miniature shielded

cable. An adapter was fabricated (lower assembly, Fig. 5) to connect the condenser to the preamplifier; the condenser was strapped to the throat of the talker by a piece of elastic, and the preamplifier was rested on the talker's lap (Fig. 7).

FIG. 7. Equipment setup for obtaining comparison between acoustic pickup and tracheal pickup.

A rubber grommet was fitted about the microphone to exclude noises, cover the side slots characteristic of this model, and form a reasonably tight coupler between the microphone and the throat over the outer trachea. The air volume of this coupler measured approximately 3 cc. This setup gave us the best results, surpassing even those given by the 640AA, with the added advantages of lightness, flexibility, and close coupling to the trachea. Figure 8 shows some samples of the tracheal signal obtained by this method of pickup with the Altec 21C microphone, where one talker uttered different sounds.

2.1.5—Attempts were also made to improvise a light comfortable unit from a small hearing-aid earpiece. This was a magnetic type, in which a stiff diaphragm behind a small port is set in motion by variations of air pressure. The voltage generated in the coils is proportional to the velocity of diaphragm displacement. The diaphragm was uncovered, and various armatures of differing stiffness and area,

FIG. 8. Fundamental periodicity derived from an Altec 21C microphone: upper, normal external acoustic signal; lower, tracheal signal without further processing.

driven by air or by direct contact, were tried. Damping materials such as sponge rubber and metallic springs were used. Fairly good response was obtained when the armature was pressed lightly against the throat, but any slight head movement had a gross effect on the output signal. The method was discarded because net results were unsatisfactory from the standpoint of frequency response.

Throughout the tests it was difficult to maintain constant coupling between the throat and the transducer. The cause appeared to be wrinkling of the skin and repositioning of the neck muscles and tissues as the head was moved. A solution was attempted by interposing a sac of fluid material between the throat and the transducer. Small soft plastic bags filled with air, water, honey, and syrup were tried, the latter two substances being readily available in individual-service restaurant packages. All these pads made a slight improvement in mobility but since they introduced excessive attenuation they were abandoned.

3. PROCESSING

Where the condenser-microphone is used either directly or coupled with the plastic tubing, very little signal-processing is required to derive clean, useful pitch pulses from even the worst speech samples (Fig. 9). Figure 10 is a block diagram of the tracheal pitch extractor. A 600-cps low-pass filter after the transducer helps suppress unwanted high-frequency components. After filtering, the signal is amplified and center-clipped by a pair of diodes. The amount of clipping is determined by the gain of the amplifier, which is variable. Properly set, this clipping removes all residual harmonic components, leaving only the fundamental pitch pulses. These pulses are passed through a simple limiter to remove any amplitude variations, and are then fed to a dual-track recorder.

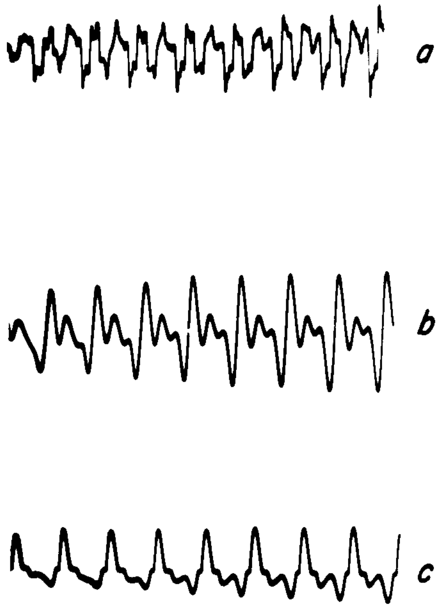

FIG. 9. Effect of performing simple processing on the tracheal signal: (a) external acoustic signal; (b) unprocessed tracheal signal; (c) tracheal signal after simple processing.

The resulting pulses (Fig. 11) indicate the basic pitch period. For many applications these pulses are adequate. For example, when these pulses are fed into a digital counter and compared with a standard time interval, an accurate short-term average of the pitch frequency is displayed. If desired, these pulses can be further refined by feeding them into a Schmitt trigger and using the leading edge of the Schmitt pulse to drive a pulse generator. This procedure gives pulses of uniform width and removes any vestige of amplitude variation that might be retained because the clipped and limited pulses have varying widths.

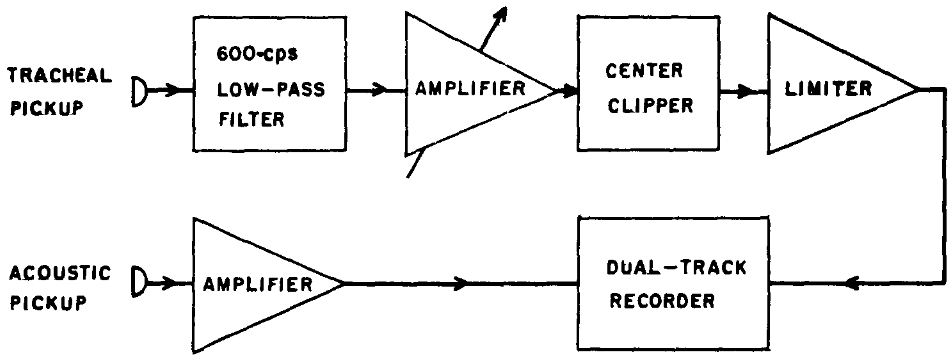

FIG. 10. Block diagram of the test setup.

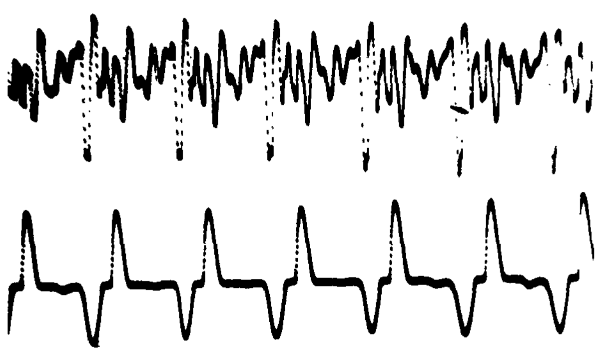

FIG. 11. Processed tracheal signal compared with external acoustic signal: upper, acoustic signal; lower, processed tracheal signal.

Prior to making recordings, the polarizations of the transducer, microphone, and all associated circuits were standardized so that a positive pressure on the transducer diaphragm gave a positive deviation in the recorded signal. An Electro-Voice 635 dynamic microphone was used for the normal voice pickup. As its output was recorded on one track of a dual-channel recorder the output of the tracheal

pickup was simultaneously recorded on the other track. With the proper phasing maintained at all times, the relationship existing between the tracheal vibrations and the acoustic output could be observed with some certainty. To further ensure a faithful picture, a minimum distance between the talker's mouth and the acoustic pickup was carefully maintained so as to reduce the time delay introduced because the vocal cord vibrations must travel through the air to the exterior pickup. The shortest reasonable air path between the larynx and external microphone was about 1.5 ft, corresponding to a frequency of approximately 650 cps. In the case of the direct pickup it was therefore impossible to avoid almost one full cycle of delay at the highest frequencies of interest. In the case of the flexible tube pickup this problem did not exist since we were able to adjust the length of the tubing to provide a delay approximating that of the voice pickup in free air.

4. DISCUSSION

Previous tests in this Laboratory have demonstrated that the direct magnetic recording method distorts waveform fidelity and that this distortion can be prevented by using FM recording techniques.[3] The samples shown in Fig. 1 were recorded on an Ampex FM recorder at 60 ips, with the 640AA microphone directly coupled to the throat and the 635 microphone situated to pick up normal speech.

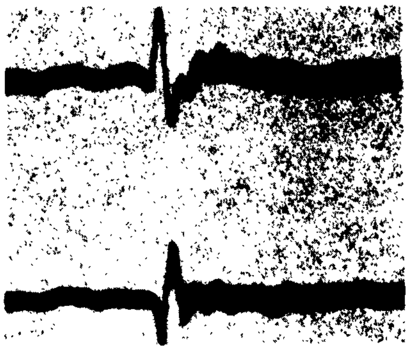

FIG. 12. Single glottal 'click' recorded and reproduced on a waveform-preserving system (FM magnetic tape): upper, tracheal pickup; lower, acoustic pickup.

The normal speech waveform and the coincident tracheal signal were recorded on dual-track tape by both direct and FM recording techniques. Several hundred feet of 35-mm film were made from the oscilloscope display of these recorded signals; in addition, Polaroid oscillophotographs were made by means of a special tape-scanning device developed in this Laboratory. The tracheal tracks of these films and tapes show a pitch periodicity that is readily apparent to even an inexperienced observer.

A simultaneous dual-track FM recording of a single "click" or vibration of the vocal cords, picked up by the condenser-microphone in contact with the trachea and by the dynamic microphone located about 1 ft from the talker's mouth, is of special interest (Fig. 12). Visual correlation of the two signals, which present an unencumbered view of what is going on inside the vocal tract both above and below the larynx, should provide a better understanding of the supra- and subglottal phenomena.

5. COMMENT

This series of experiments has provided us with several valuable results. By any standard that we have applied, the system appears to constitute an accurate, reliable electronic pitch extractor. In several hundred feet of film, we have discovered no abrupt discontinuity in periodicity. Where the external speech signal is sufficiently strong to be observed, an excellent visual correlation exists between the tracheal pulses and those aspects of the speech wave commonly associated with voicing excitation. Careful comparison of the two waveforms indicates complete correlation, although an absolute numerical count has not yet been attempted. In addition, the tracheal pickup shows glottal excitations at instants when little or no external speech signal is discernible, for example, during closure in voiced stops and during nasals. To our knowledge, in the absence of more precise indication of vocal cord activity such as might be obtained through direct visual observation, this highly reliable technique for pitch extraction is certainly the most accurate electronic method yet devised.

References

1. TOSHITAKA SUGIMOTO and SHIZUO HIKI, Extraction of the pitch of a voice from the vibration of the outer skin of the trachea, J. Acoust. Soc. Japan **16**(No. 4):291-293, Dec. 1960.
2. ———, On the extraction of the pitch signal using the body wall vibration at the throat of the talker, Paper G26, IVth Intl Congress on Acoustics, Copenhagen, Denmark, 21 to 28 August 1962.
3. ROY L. KOMACK, Waveform Preservation in Magnetic Tape Recording, AFCRL-ERD-CRRSV-7M-61-8, Air Force Cambridge Research Center, December 1961.

AF Cambridge Research Laboratories, Bedford, Mass. Electronics Research Directorate
EXTRACTION OF PITCH FROM THE TRACHEA
by Harold C. Porter. 10 pp. incl. illus.
February 1963. AFCRL-63-24.
Unclassified report

For decades, speech researchers have recognized that accurate pitch information is prerequisite to the synthesis or reconstitution of speech that has been broken into its basic parts for any reason, such as communication over restricted bandwidth media. In a series of experiments investigating tracheal pitch extraction according to a methodology suggested in a Japanese paper, the Speech Research Branch, AFCRL Communication Sciences Laboratory succeeded in achieving reasonably accurate real-time pitch extraction. The method has a limitation in that the talker must wear a special appliance.

UNCLASSIFIED

1. Pitch extraction, Tracheal

I. Porter, Harold C.

UNCLASSIFIED

AF Cambridge Research Laboratories, Bedford, Mass. Electronics Research Directorate
EXTRACTION OF PITCH FROM THE TRACHEA
by Harold C. Porter. 10 pp. incl. illus.
February 1963. AFCRL-63-24.
Unclassified report

For decades, speech researchers have recognized that accurate pitch information is prerequisite to the synthesis or reconstitution of speech that has been broken into its basic parts for any reason, such as communication over restricted bandwidth media. In a series of experiments investigating tracheal pitch extraction according to a methodology suggested in a Japanese paper, the Speech Research Branch, AFCRL Communication Sciences Laboratory succeeded in achieving reasonably accurate real-time pitch extraction. The method has a limitation in that the talker must wear a special appliance.

UNCLASSIFIED

1. Pitch extraction, Tracheal

I. Porter, Harold C.

UNCLASSIFIED

AF Cambridge Research Laboratories, Bedford, Mass. Electronics Research Directorate
EXTRACTION OF PITCH FROM THE TRACHEA
by Harold C. Porter. 10 pp. incl. illus.
February 1963. AFCRL-63-24.
Unclassified report

For decades, speech researchers have recognized that accurate pitch information is prerequisite to the synthesis or reconstitution of speech that has been broken into its basic parts for any reason, such as communication over restricted bandwidth media. In a series of experiments investigating tracheal pitch extraction according to a methodology suggested in a Japanese paper, the Speech Research Branch, AFCRL Communication Sciences Laboratory succeeded in achieving reasonably accurate real-time pitch extraction. The method has a limitation in that the talker must wear a special appliance.

UNCLASSIFIED

1. Pitch extraction, Tracheal

I. Porter, Harold C.

UNCLASSIFIED

AF Cambridge Research Laboratories, Bedford, Mass. Electronics Research Directorate
EXTRACTION OF PITCH FROM THE TRACHEA
by Harold C. Porter. 10 pp. incl. illus.
February 1963. AFCRL-63-24.
Unclassified report

For decades, speech researchers have recognized that accurate pitch information is prerequisite to the synthesis or reconstitution of speech that has been broken into its basic parts for any reason, such as communication over restricted bandwidth media. In a series of experiments investigating tracheal pitch extraction according to a methodology suggested in a Japanese paper, the Speech Research Branch, AFCRL Communication Sciences Laboratory succeeded in achieving reasonably accurate real-time pitch extraction. The method has a limitation in that the talker must wear a special appliance.

UNCLASSIFIED

1. Pitch extraction, Tracheal

I. Porter, Harold C.

UNCLASSIFIED